Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes†

Chenping Xu,[1] Thomas J Caperna,[2] Wesley M Garrett,[3] Perry Cregan,[4] Hanhong Bae,[5] Devanand L Luthria[6] and Savithiry Natarajan[4,*]

[1]*University of Maryland, Department of Plant Science and Landscape Architecture, College Park, MD 20742, USA*
[2]*U.S. Department of Agriculture, Agricultural Research Service, Growth Biology Laboratory, Beltsville, MD 20705, USA*
[3]*U.S. Department of Agriculture, Agricultural Research Service, Biotechnology and Germplasm Laboratory, Beltsville, MD 20705, USA*
[4]*U.S. Department of Agriculture, Agricultural Research Service, Soybean Genomics and Improvement Laboratory, PSI, Beltsville, MD 20705, USA*
[5]*U.S. Department of Agriculture, Agricultural Research Service, Sustainable Perennial Crops Laboratory, Beltsville, MD 20705, USA*
[6]*U.S. Department of Agriculture, Agricultural Research Service, Food Composition Laboratory, Beltsville, MD 20705, USA*

Abstract

BACKGROUND: A comparative analysis of seed allergens from various soybean genotypes is crucial for identifying and eliminating potential allergens. We have investigated the distribution of three major allergens (Gly m Bd 60K, Gly m Bd 30K and Gly m Bd 28K) in wild, landrace, ancestral and modern soybean genotypes.

RESULTS: Gly m Bd 60K allergens consist of α subunits of β-conglycinin and G2 subunits of glycinin. In wild genotypes, α subunits of β-conglycinin separated into six to seven protein spots whereas five to seven spots were observed in the landraces. All genotypes of modern and ancestral groups showed 3–5 protein spots of α subunits of β-conglycinin. All genotypes showed eight spots of glycinin G2 subunits except one ancestral genotype which had seven spots. Two protein spots were detected for Gly m Bd 30K in 14 genotypes but one spot was detected in two wild genotypes. Two protein spots were detected for Gly m Bd 28K in all genotypes.

CONCLUSION: Considerable heterogeneity of the α subunit of β-conglycinin distribution exists among these 16 soybean genotypes. Significant proteomic variation was observed between different soybean groups rather than among genotypes in the same group. This investigation would be valuable to researchers working with soybean and nutrition.

Keywords: 2D-PAGE; soybean; proteomics; allergen; MALDI-TOF-MS; LC-MS/MS

INTRODUCTION

Seed legumes provide one-fifth of all plant proteins consumed by humans on a global basis.[1] Soybeans (*Glycine max* [L.] Merr.) have long been recognized as a valuable source of edible oil and protein for feeding both animals and humans.[2] However, it ranks among the eight most significant food allergens, to which more than 90% of all food allergies in the United States are attributed.[3] In the USA and Europe, between 5 and 8% of babies and 2% of adults are reported to be allergic to soybeans.[4] Increased awareness of several health benefits associated with the consumption of soybean protein has resulted in widespread use of a variety of food products manufactured with soybeans. Therefore, it is very difficult for sensitive individuals to avoid soy products in prepared and processed foods.[5,6]

Soybean possesses about 15 proteins recognized by IgEs from soy-sensitive people.[7,8] Three proteins, Gly m Bd 60 K, Gly m Bd 30 K, and Gly m Bd 28 K, represent the major seed allergens in soybean-sensitive patients.[9] Gly m Bd 60 K allergenic proteins are grouped into two types: β-conglycinin and glycinin. Only the α subunit of β-conglycinin, one of the very abundant soybean seed storage proteins,[10] has allergenic reactions in about 25% of sera from soybean sensitive patients.[11] Glycinin, another storage protein

* Correspondence to: Savithiry Natarajan, Soybean Genomics and Improvement Laboratory, PSI-ARS-USDA, 10300 Baltimore Avenue, Beltsville, MD 20705, USA
E-mail: natarajs@ba.ars.usda.gov
† Mention of a trade name, proprietary product or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable.
Contract/grant sponsor: ARS project; contract/grant number: 1275-21000-156-00D
(*Received 8 February 2007; revised version received 1 May 2007; accepted 6 May 2007*)
Published online 13 August 2007; DOI: 10.1002/jsfa.3017

and the most prevalent protein in soybean seed, consists of five subunits, G1, G2, G3, G4 and G5. Among these subunits, only G2 and acidic polypeptides of G1 are reported to be allergens.[12] Gly m Bd 30 K, a member of the papain superfamily of cysteine proteases, also referred as P34, has been identified as a major allergen in soybean seeds, because more than 65% of soy-sensitive patients react to the Gly m Bd 30 K protein.[8,13,14] Among all cysteine proteases, P34 remains unique because of its active site modification and unknown function in soybean.[15] P34 has a tendency to bind to lipid and oil, which is likely significant in its role as an allergen and as a possible defense protein.[16] Gly m Bd 28 K is an MP27–MP33 homolog, a minor soybean seed globulin.[6] It belongs to the cupin protein superfamily, which contains a number of legume food allergens, such as the peanut Ara h 1 and Ara h 3 as well as soybean β-conglycinin and glycinin G1 and G2 proteins.[17]

Many soybean varieties have been developed for adaptation to various geographical regions via plant breeding for increasing the yield of the crop and quality of seeds. In general, protein composition and profiles can be affected by breeding and environmental conditions.[18,19] Furthermore, comparative proteomic studies of different soybean allergens would aid in the understanding of evolutionary relationships between these genotypes. Therefore, we have conducted comparative studies of allergen proteins in sixteen soybean genotypes including cultivated and wild soybean genotypes. Cultivated soybeans are derived from wild soybeans (*Glycine soja*). The cultivated soybeans can be crossed with *G. soja* to produce fertile offspring. Wild species are sources of genes for crop improvement.[20]

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) can be used to attain a comprehensive protein profile. This technique is known as an effective method to simultaneously resolve a large number of proteins.[21,22] When associated with mass spectrometry (MS), it can allow the identification and characterization of selected proteins.[23] Combined proteomic technologies are very powerful for accurately examining soybean seed protein composition.[24,25] Using proteomic technologies, we have already successfully compared the profiles of anti-nutritional proteins (Kunitz trypsin inhibitors)[26] and storage proteins[27] in soybean seed among different genotypes. In this study, we continued to characterize and compare the distributions of the three major allergenic proteins (Gly m Bd 60 K, Gly m Bd 30 K and Gly m Bd 28 K) among 16 genotypes. A comparative analysis of seed allergens from various soybean genotypes is crucial for identifying and eliminating potential soybean allergens. Such evaluations will be useful for comparative analysis of the allergenicity of other legume species. The three major allergens were well-separated using 2D-PAGE, and identified by matrix-assisted laser desorption/ionization time-of-flight MS (MALDI–TOF-MS) and liquid chromatography MS (LC–MS/MS). This investigation on variation of allergens among 16 soybean genotypes will enable us to have a better understanding of the natural variation of seed allergens.

MATERIALS AND METHODS

Plant materials

Soybean seeds of wild soybean (*G. soja*) genotypes (PI 407 027, PI 407 282, PI 366 120 and PI 393 551), Asian landrace genotypes (PI 423 954, PI 89 138, PI 594 777 and PI 59 845), ancestral genotypes of North American soybean (PI 548 298 (AK-Harrow), PI 548 445 (CNS), PI 548 318 (Dunfield) and PI 548 362 (Lincoln)), and modern cultivars (PI 533 655 (Burlison), PI 525 453 (Conrad), PI 513 382 (Glenwood) and PI 536 635 (Sprite)) were obtained from the USDA soybean germplasm collection (Urbana, IL, USA). All seeds were stored at −80 °C until analyzed.

Chemicals

Chemicals for electrophoresis, including acrylamide, bis-acrylamide, sodium dodecyl sulfate (SDS), tetramethylethylenediamine (TEMED), ammonium persulfate, thiourea, dithiothreitol (DTT), 3-[(3-cholamidopropyl) dimethylamonio]-1-propane-sulfonate (CHAPS) and immobilized pH gradient (IPG) strips were purchased from GE Healthcare (Piscataway, NJ, USA). Urea and ampholytes (pH 4.0–7.0 and 3.0–10.0) were purchased from Bio-Rad Laboratories (Hercules, CA, USA). Tris-HCl (pH 8.8), 2-mercaptoethanol, trichloroacetic acid (TCA) and glycerol were purchased from Sigma-Aldrich Chemical Co. (St Louis, MO, USA). α-Cyanohydroxycinnamic acid (CHCA) matrix was purchased from Bruker Daltonics (Billerica, MA, USA). Water from a Millipore Milli-RO4 reverse osmosis system (Millipore, Newbedford, MA, USA) was used for making all solutions.

Extraction and 2D-PAGE analysis

Protein extraction from seed and 2D-PAGE separation protocols were performed according to the procedures described in our earlier publication.[25] One hundred milligrams of the soybean seed powder was homogenized with solution containing 10% (w/v) TCA in acetone with 0.07% (v/v) 2-mercaptoethanol. The protein was precipitated for 1 h at −20 °C. The extract was centrifuged (Eppendorf 5804 R; Brinkman Instruments Inc., Westbury, NY, USA) and the pellet was washed with acetone containing 0.07% (v/v) 2- mercaptoethanol. Then, the pellet was dried under vacuum and resuspended in 1 mL of lysis buffer [9 mol L- urea, 1% CHAPS, 1% (w/v) one of the ampholytes (pH 4.0–7.0, and 3.0–10.0)] and 1% DTT followed by sonication for 30 min. Insoluble material was removed by centrifugation, and the supernatant was used for 2D-PAGE analysis. The first-dimension isoelectric focusing (IEF) was performed using 13 cm pH

J Sci Food Agric **87**:2511–2518 (2007)
DOI: 10.1002/jsfa

4.0–7.0 or 3.0–10.0 linear IPG strips in an IPGphor apparatus (GE Healthcare, Piscataway, NJ, USA), according to the manufacturer's recommendations. For the second dimension, the IPG strips were incubated with 50 mmol L^{-1} Tris-HCl, pH 8.8, 6 mol L^{-1} urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, and 1% DTT for 15 min, acetylated with iodoacetamide, and subsequently placed onto 12% polyacrylamide gels. Electrophoresis was performed using a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare, Piscataway, NJ, USA): 5 mA gel^{-1} for 30 min in the first step, then 17 mA gel^{-1} for 5 h in the second step. The 2D-PAGE gels were visualized by staining with colloidal Coomassie blue G-250, and scanned using laser densitometry (GE Healthcare, Piscataway, NJ, USA). Triplicate samples were used for soybean seed protein extraction and 2D-PAGE analysis.

In-gel digestion of protein spots

Protein spots were excised from the stained gel and washed first with distilled water to remove ammonium sulfate and then with 50% acetonitrile containing 25 mmol L^{-1} ammonium bicarbonate to destain the gel plug. The gel plug was dehydrated with 100% acetonitrile, dried under vacuum, and then re-swollen with 20 μL of 10 μg mL^{-1} trypsin (modified porcine trypsin, sequencing grade, Promega, Madison, WI) in 25 mmol L^{-1} ammonium bicarbonate. Digestion was performed overnight at 37 °C. The resulting tryptic fragments were extracted with 50% acetonitrile and 5% trifluoroacetic acid with sonication. The extract was dried to completeness and dissolved in 50% acetonitrile and 0.1% TCA.

Mass spectrometry

For peptide mass fingerprinting (PMF), a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems, Framingham, MA, USA) operated in positive ion reflector mode was used to analyze tryptic peptides. Samples were co-crystallized with CHCA matrix, and spectra were acquired with 50 shots of a 337 nm nitrogen laser operating at 20 Hz. Spectra were calibrated using the trypsin autolysis peaks at m/z 842.51 and 2211.10 as internal standards. For LC–MS/MS, a Thermo Finnigan LCQ Deca XP plus ion trap mass spectrometer was used to analyze proteins that were not positively identified by MALDI–TOF-MS. Peptides were separated on a reverse phase column using a 30 min gradient of 5–60% acetonitrile in water with 0.1% formic acid. The instrument was operated with a duty cycle that acquired MS/MS spectra on the three most abundant ions identified by a survey scan from 300 to 2000 Da. Dynamic exclusion was employed to prevent the continuous analysis of the same ions. Once two MS/MS spectra of any given ion had been acquired, the parent mass was placed on an exclusion list for the duration of 1.5 min. The raw data were processed by Sequest to generate DTA files for database searching. The merge pl script from Matrix Science was used to convert multiple Sequest DTA files into a single mascot generic file suitable for searching in Mascot.

Protein identification

Protein identification was performed using the Mascot search engine. NCBInr database was selected as the primary database to be searched. The parameters for database searches with MALDI–TOF PMF data and with MS/MS spectra were set as before.[25] For MALDI–TOF-MS data to qualify as a positive identification, a protein's molecular weight search (MOWSE) score had to equal or exceed the minimum significant score of 64. Positive identifications of proteins by MS/MS analysis required a minimum of two unique peptides, with at least one peptide having a significant ion score.

RESULTS AND DISCUSSION

In this study, we have separated major soybean allergens in 16 soybean genotypes (four each of wild soybean, Asian landraces, North American ancestors and modern cultivars) using 2D-PAGE. We analyzed proteins with two pH ranges (3.0–10.0 and 4.0–7.0), because the G2 subunit of glycinin was not separated well in pH range 3.0–10.0. These protein spots were identified by MALDI–TOF-MS and LC–MS/MS. We primarily focused on the distribution, comparison and variation of the three major allergens, Gly m Bd 60 K, Gly m Bd 30 K, and Gly m Bd 28 K in the soybean seed. Table 1 includes the origin, cultivar types and strain designation of all the 16 genotypes. Figures 1 and 2 show representative pH 3.0–10.0 and 4.0–7.0 gel images of one wild (PI 407 027) and one cultivated (landrace, PI 59 845) soybean seed sample. Table 2 lists the spots identified as major allergens in sixteen soybean genotypes. In Table 3, spot numbers of allergens observed in each genotype are listed. Our results show that the overall distribution of the allergens is similar in all four groups of genotypes. However, the number of spots and their intensities varied among these genotypes (Table 3; Figs 1 and 2).

Analysis of Gly m Bd 60 K proteins

Soybean Gly m Bd 60 K allergens include β-conglycinin and glycinin. Only the α subunit of β-conglycinin has allergenic reactions in about 25% of sera from soybean sensitive patients.[11] G2 subunits and acidic polypeptides of G1 subunits of glycinin are reported to be allergenic.[12] All glycinin subunits are composed of acidic and basic polypeptides. The acidic polypeptides of the G1 subunit were not observed in this study, probably because these proteins are too low in abundance to be detected by this method. The basic polypeptides were observed, but the results are not presented because they do not elicit an allergic reaction.[12] Considerable variation was observed for the α subunit of β-conglycinin. In our study, all the wild and landrace genotypes showed five spots (spots 1, 3, 4, 6 and 7 of the α subunit). One wild (PI

Table 1. The 16 soybean genotypes tested for protein variation

Number	Soybean group	Cultivar	Strain designation	Origin: Province or state/country
1	Wild (*G. soja*)	n/a	PI 366120	Akita (Japan)
2	Wild (*G. soja*)	n/a	PI 393551	Taiwan (Taiwan)
3	Wild (*G. soja*)	n/a	PI 407027	Akita (Japan)
4	Wild (*G. soja*)	n/a	PI 407282	Cheju (South Korea)
5	Landrace (*G. max*)	Shirome	PI 423954	Kumamota (Japan)
6	Landrace (*G. max*)	Zontanorukon	PI 89138	Hamgyong Puk (North Korea)
7	Landrace (*G. max*)	Liuyue huang	PI 594777	Yunnan (China)
8	Landrace (*G. max*)	Sohgetsu	PI 59845	Japan
9	Ancestor (Old)	A.K (Harrow)	PI 548298	China
10	Ancestor (Old)	CNS	PI 548445	Jiangsu (China)
11	Ancestor (Old)	Dunfield	PI 548318	Jilin (China)
12	Ancestor (Old)	Lincoln	PI 548362	Pedigree unknown (USA)
13	Modern (Elite)	Burlison	PI 533655	Illinois (USA)
14	Modern (Elite)	Conrad	PI 525453	Iowa (USA)
15	Modern (Elite)	Glenwood	PI 513382	Minnesota (USA)
16	Modern (Elite)	Sprite	PI 536635	Ohio (USA)

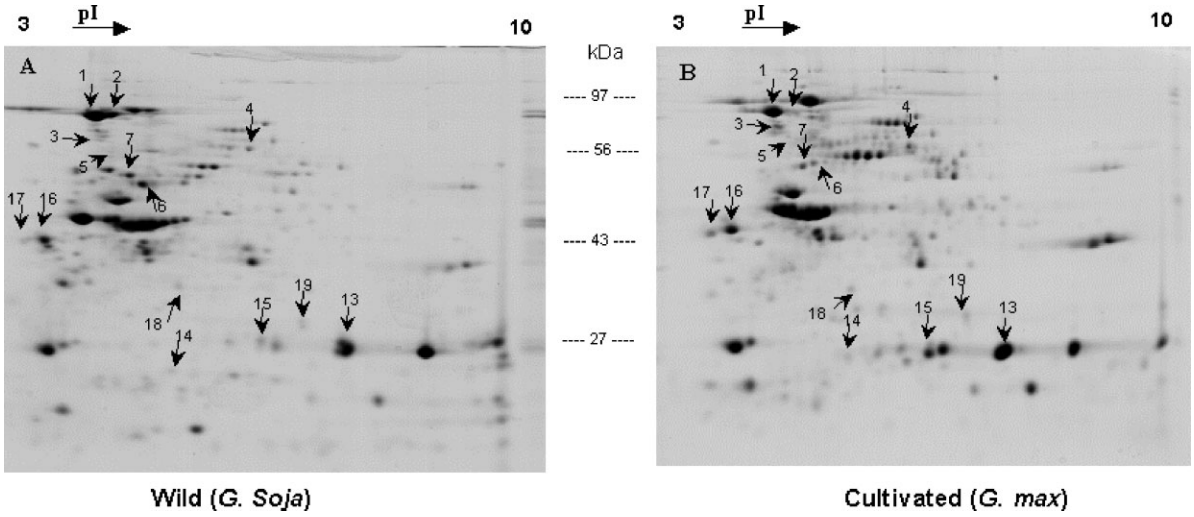

Figure 1. A proteomic comparison of the allergens of wild, *G. soja*, PI 407027 (1A) and landrace, *G. max*, PI 59845 (1B) soybean seeds. The first dimension was run using a pH gradient from 3.0 to 10.0 and the second dimension was a 12% SDS-PAGE. Gels were stained with colloidal Coomassie blue stain G-250. Numbered arrows indicate the spots referred to in the text and tables.

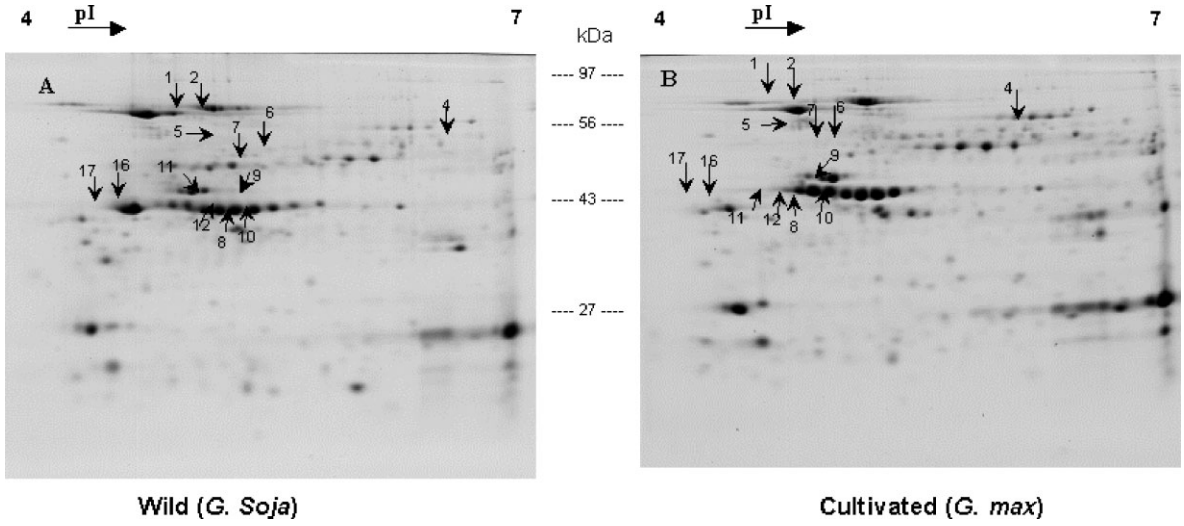

Figure 2. A proteomic comparison of the allergens of wild, *G. soja*, PI 407027 (A) and landrace, *G. max*, PI 59845 (B) soybean seeds. The first dimension was run using a pH gradient from 4.0 to 7.0 and the second dimension is a 12% SDS-PAGE. Gels were stained with colloidal Coomassie blue stain G-250. Numbered arrows indicate the spots referred to in the text and tables.

J Sci Food Agric **87**:2511–2518 (2007)
DOI: 10.1002/jsfa

Table 2. Major allergenic proteins identified by MS in soybean seeds

Sl #	Theoretical pI/mol. wt	Protein identity	Identification method	Peptides matched	Sequence coverage (%)	MOWSE score	NCBI accession #
1	4.92/63 127	α subunit of β-conglycinin	MALDI–TOF-MS	25	39	217	gi|9967357
2	4.92/63 127	α subunit of β-conglycinin	MALDI–TOF-MS	25	41	167	gi|9967357
3	4.92/63 184	α subunit of β-conglycinin	MALDI–TOF-MS	27	43	250	gi|9967357
4	4.92/63 184	α subunit of β-conglycinin	MALDI–TOF-MS	20	34	204	gi|9967357
5	5.32/72 717	α subunit of β-conglycinin	MALDI–TOF-MS	9	18	73	gi|15425633
6	5.32/72 717	α subunit of β-conglycinin	MALDI–TOF-MS	14	24	112	gi|15425633
7	5.32/72 717	α subunit of β-conglycinin	LC–MS/MS	9	23	238	gi|51247837
8	5.46/54 927	Glycinin G2 precursor	MALDI–TOF-MS	9	19	91	gi|1212177
9	5.46/54 927	Glycinin G2 precursor	MALDI–TOF-MS	9	19	73	gi|1212177
10	5.46/54 927	Glycinin G2 precursor	MALDI–TOF-MS	8	15	71	gi|1212177
11	5.46/54 927	Glycinin G2 precursor	MALDI–TOF-MS	9	14	74	gi|1212177
12	5.46/54 927	Glycinin G2 precursor	MALDI–TOF-MS	12	21	104	gi|1212177
13	5.78/54 047	Glycinin G2 precursor	MALDI–TOF-MS	6	37	67	gi|169967
14	5.56/54 903	Glycinin G2 precursor	LC–MS/MS	3	9	175	gi|72295
15	5.56/54 903	Glycinin G2 precursor	LC–MS/MS	9	14	313	gi|72295
16	5.74/43 136	P34 probable thiol protease precursor	LC–MS/MS	4	9	102	gi|129353
17	5.74/43 136	P34 probable thiol protease precursor	LC–MS/MS	2	9	60	gi|129353
18	5.73/52 813	Allergen Gly m Bd 28 K	LC–MS/MS	3	6	187	gi|12697782
19	5.73/52 780	Allergen Gly m Bd 28 K	MALDI–TOF-MS	10	19	97	gi|12697782

Table 3. Soybean allergen composition in 16 genotypes

		Protein spot number of allergenic proteins			
		Gly m Bd 60 K			
Soybean group	Accession	α subunit of β-conglycinin	Glycinin G2	Gly m Bd 30 K	Gly m Bd 28 K
Wild (*G. soja*)	PI366120	1, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16	18, 19
	PI393551	1, 2, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16	18, 19
	PI407027	1, 2, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI407282	1, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
Landrace (*G. max*)	PI423954	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI89138	1, 2, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI594777	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI59845	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
Ancestor (old)	PI548445	1, 3, 4	8, 9, 10, 11, 12, 13, 14	16, 17	18, 19
	PI548298	1, 3, 4	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI548318	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI548362	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
Modern (elite)	PI536635	1, 3, 4	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI533655	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI525453	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI513382	1, 3, 4	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19

393 551) and one landrace (PI 89 138) genotype show two additional spots (spots 2 and 5), and two wild genotypes (PI 366 120 and PI 407 282) have one extra spot (spot 5).

All ancestral and modern genotypes have spots 1, 3 and 4 of the α subunit, while two ancestors (PI 548 318 and PI 548 362) and two modern cultivars (PI 533 655 and PI 525 453) show two additional spots (spots 6 and 7). The glycinin allergens showed eight spots (spots 8–15) and are identified as G2 subunits (Table 2). The distribution of G2 subunits was very similar among these genotypes. Fifteen genotypes showed eight spots (spots 8–15), but only one ancestral genotype (PI 548 445) showed seven spots of G2 subunits (spots 8–15), with spot 15 missing (Table 3).

The distribution of Gly m Bd 60 K (β-conglycinin and glycinin) in cultivated soybeans has been reported previously.[24,28] Little information is available about the variation of Gly m Bd 60 K allergens among different soybean genotypes. A mutant line without the α subunit of β-conglycinin has been developed by irradiation.[29–31] However, the protein from this mutant retains no gel-forming ability for use in traditional products and no data were available as to the nutritional quality of this soybean.[9] In another study, the expression of α and α' subunits of

β-conglycinin was suppressed by sequence-mediated gene silencing in transgenic soybean seed.[32] In this case, the decrease in β-conglycinin protein was apparently compensated by an increased accumulation of glycinin and P34. Takahashi *et al*.[33] have generated a soybean mutant line, QF2, whose seeds lack both glycinin and β-conglycinin. This mutant line grew and reproduced normally and the nitrogen content of its dry seed was similar to the wild type cultivars. However, this line exhibits an abundance of other proteins including allergen protein Gly m bd 30 K (P34), the immunodominant allergen and other anti-nutritional protein such as lectin and lipoxygenase.

In our study, multiple spots are observed for the α subunit of β-conglycinin, acidic and basic polypeptides of the G2 subunit. The migration of proteins on a 2D-PAGE gel is very sensitive to small structural differences. One gene product may undergo different co- and/or post-translational modifications that affect its pI or/and molecular weight, and it is known that the α subunit of β-conglycinin undergoes extensive co- and post-translational modification.[34]

Analysis of Gly m Bd 30 K

Gly m Bd 30 K is a major allergen protein and has been shown to be identical to the previously described soybean seed 34 kDa vacuolar protein, P34.[8] P34 has been shown to account for a majority of IgE binding with serum from several soy-sensitive humans.[35] It is a relatively minor seed constituent, comprising less than 1% of total seed protein. The P34 protein can be detected in a wide variety of soybean processed foods.[6] The elimination of P34 from soybean seeds would enhance food safety and make the use of soybean products available to sensitive individuals. Using a monoclonal antibody against P34, Yaklich *et al*.[35] found that P34 was almost uniformly distributed in the genetic diversity of soybean. The high protein lines did not contain more P34 than other lines, and they did not find naturally occurring varieties lacking P34 allergen protein, when they surveyed the core collection representing the public cultivars released between 1947 and 1988 in North America.[35] A similar survey of Japanese lines did not reveal any P34 null lines.[36] Consistent with these results, this study shows that P34 is present in all these sixteen soybean genotypes examined (Table 3). However, two wild soybean genotypes (PI 366 120 and PI 393 551) show only one spot (spot 16) while the other 14 soybean genotypes show two spots of P34 (spots 16 and 17). P34 spot(s) show weak intensity in the wild genotype compared to that in the landrace genotype (Fig. 2A and B). Recently, Joseph *et al*.[37] screened 16 266 accessions of USDA soybean germplasm, and identified 12 genotypes as P34 null lines. Such varieties of an allergen null/reduced soybean could be especially valuable for uses in baby formula. Biotechnology offers the prospect of eliminating P34 allergenic protein from soybean. Herman[15] and Herman *et al*.[28] have used gene silencing to successfully eliminate accumulation of P34 in transgenic soybean.

Analysis of Gly m Bd 28 K protein

Gly m Bd 28 K is a less abundant protein of soybean but is designated as a major allergen, and was originally isolated from soybean meal as a 28 kDa glycosylated protein.[38] In this study, two spots (spots 18 and 19) of Gly m Bd 28 K were identified by LC–MS/MS in all 16 genotypes (Table 2). No major variation of this protein is observed in these genotypes (Table 3). There is very little additional information about variation in this allergen among different soybean genotypes. A mutant line of soybean *G. max*, cv Tohoku 124, with the absence of β-conglycinin also lacks Gly m Bd 28 K.[30]

Xiang *et al*.[39] demonstrated that both the N- and C-terminal fragments of Gly m Bd 28 K can bind IgE, and have characterized a novel dominant sequential epitope on the C-terminal fragment of the protein. Tsuji *et al*.[40] reported that Gly m Bd 28 K was about twice as large as anticipated from its initial discovery as a 28 kDa polypeptide by one-dimensional SDS-PAGE. The Gly m Bd 28 K allergen is probably processed into two smaller polypeptides of 240 and 212 amino acids in the soybean seed, and both portions are expected to be present in soybean-derived foods.[39,40] The full-length Gly m Bd 28 K is encoded by an open reading frame of 473 amino acids.[39] Consistent with these reports, the amino acid sequence of spot 18 from our study only matches the second half of the protein, and that of spot 19 only matches the first half. These two spots of Gly m Bd 28 K probably come from the post-translational processing of the same gene.

CONCLUSION

In summary, considerable heterogeneity of the α subunit of β-conglycinin protein spot distribution exists among these 16 soybean genotypes. Major proteomic variation was observed between soybean groups rather than among genotypes in the same group, which might be due to the breeder's efforts to select better genotypes for protein production, while the other allergen Gly m Bd 60 K (G2 subunit of glycinin) distributes similarly among these 16 genotypes. Fourteen genotypes show two protein spots and two genotypes show one spot of allergen Gly m Bd 30 K. All 16 genotypes show two spots of Gly m Bd 28 K, and no variation in protein pattern is observed. The data from this study may serve as criteria for profiling and the comparison of soybean allergenic proteins in various soybean genotypes and also useful for the modification of soybeans to improve its nutrition values.

J Sci Food Agric **87**:2511–2518 (2007)
DOI: 10.1002/jsfa

ACKNOWLEDGEMENTS

We thank Dr J. Slovin for critical review of this manuscript. Funding for this research was provided by ARS project 1275-21000-156-00D.

REFERENCES

1 Pinstrup-Andersen P, Pandya-Lorch R and Rosegrant MW, World food prospects: critical issues for the early 21st century, in *Food Policy Report*. International Food Policy Research Institute, Washington, DC, pp. 5–32 (1999).
2 Liener IE, Implications of antinutritional components in soybean foods. *Crit Rev Food Sci Nutr* **34**:31–67 (1994).
3 Taylor SL and Hefle SL, Will genetically modified foods be allergenic? *J Allergy Clin Immunol* **107**:765–771 (2001).
4 Heppel LM, Sissons JW and Pedersen HE, A comparison of the antigenicity of soya-bean-based infant formulas. *Br J Nutr* **58**:393–403 (1987).
5 Herian AM, Taylor SL and Bush RK, Allergenicity of various soybean products as determined by RAST inhibition. *J Food Sci* **58**:385–390 (1993).
6 Tsuji HN, Okada R, Yamanishi N, Bando M, Kimoto M and Ogawa T, Measurement of Gly m Bd 30K, a major soybean allergen, in soybean products by a sandwich enzyme-linked immunosorbent assay. *Biosci Biotech Biochem* **59**:150–151 (1995).
7 Burks AW, Castell HB, Fiedorek SC, Williams LW, Connaughton C and Brooks JB, Enzyme-linked immunosorbant assay and immunoblotting determination of antibody response to major component proteins of soybeans in patients with soy protein intolerance. *J Pediatr Gastroenterol Nutr* **8**:195–203 (1989).
8 Ogawa T, Tsuji H, Bando N, Kitamura K, Zhu YZ, Hirano H, *et al*, Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci Biotechnol Biochem* **57**:1030–1033 (1993).
9 Ogawa T, Samoto M and Takahashi K, Soybean allergens and hypoallergenic soybean products. *J Nutr Sci Vitaminol* **46**:271–279 (2000).
10 Ogawa T, Bando N, Tsuji H, Nishikawa K and Kitamura K, Alpha-subunit of beta-conglycinin, an allergenic protein recognized by IgE antibodies of soybean-sensitive patients with atopic dermatitis. *Biosci Biotechnol Biochem* **59**:831–833 (1995).
11 Wilson S, Blaschek K, Gonzalez E and Mejia D, Allergenic proteins in soybean: processing and reduction of P34 allergenicity. *Nutr Rev* **63**:47–58 (2005).
12 Beardslee TA, Zeece MG, Sarath G and Markwell JP, Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *Int Arch Allergy Immunol* **123**:299–307 (2000).
13 Helm RM, Cockrell G, West CM, Herman EM, Sampson HA, Bannon GA, *et al*, Mutational analysis of the IgE-binding epitopes of P34/Gly m1. *J Allergy Clin Immunol* **105**:378–384 (2000).
14 Helm RM, Cockrell G, Herman EM, Burks AW, Sampson HA and Bannon GA, Cellular and molecular characterization of a major soybean allergen. *Int Arch Allergy Immunol* **117**:29–37 (1998).
15 Herman EM, Soybean allergenicity and suppression of the immunodominant allergen. *Crop Sci* **45**:462–467 (2005).
16 Okinaka Y, Yang CH, Herman EM, Kinney AJ and Keen NT, The P34 syringolide elicitor receptor interacts with a soybean photorespiration enzyme, NADH-dependent hydroxypyruvate reductase. *Plant Mol Microb Interact* **15**:1213–1218 (2002).
17 Mills EN, Jenkins J, Marigheto N, Belton PS, Gunning AP and Morris VJ, Allergens of the cupin superfamily, *Biochem Soc Trans* **30**:925–929 (2002).
18 Burton JW, Breeding soybean cultivars for increased soybean percentage, in *Proc. World Soybean Research Conf. IV*, ed. by Pascal AJ. Association Argentina de la Soja, Buenos Aires, pp. 1079–1089 (1989).
19 Cui Z, Carter TE, Burton JW and Wells R, Phenotypic diversity of modern Chinese and North American soybean cultivars. *Crop Sci* **41**:1954–1967 (2001).
20 Jauhar PP, Modern biotechnology as an integral supplement to conventional plant breeding: the prospects and challenges. *Crop Sci* **46**:1841–1859 (2006).
21 Watson BS, Asirvatham VS, Wang L and Sumner LW, Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiol* **131**:1104–1123 (2003).
22 Komatsu S, Kojima K, Suzuki K, Ozaki K and Higo K, Rice Proteome Database based on two-dimensional polyacrylamide gel electrophoresis: its status in 2003. *Nucleic Acids Res* **32**:D388–D392 (2004).
23 Aebersold R and Mann M, Mass spectrometry-based proteomics. *Nature* **422**:198–207 (2003).
24 Mooney BP and Thelen JJ, High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochem* **65**:1733–1744 (2004).
25 Natarajan S, Xu C, Caperna TJ and Garrett WM, Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal Biochem* **342**:214–220 (2005).
26 Natarajan S, Xu C, Bae H and Bailey BA, Proteomic and genomic characterization of Kunitz trypsin inhibitors in wild and cultivated soybean genotypes. *J Plant Physiol* **164**:756–763 (2007).
27 Natarajan S, Xu C, Bae H, Caperna TJ and Garrett WM, Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *J Agric Food Chem* **54**:3114–3120 (2006).
28 Herman EM, Helm R, Jung R and Kinney A, Genetic modification removes an immunodominant soybean allergen. *Plant Physiol* **132**:36–43 (2003).
29 Hayashi M, Harada K, Fujiwara T and Kitamura K, Characterization of a 7S globulin-deficient mutant of soybean (*Glycine max* (L.) Merrill). *Mol Gen Genet* **258**:208–214 (1998).
30 Samoto M, Fukuda Y, Takahashi K, Tabuchi K, Hiemori M, Tsuji H, *et al*, Substantially complete removal of three major allergenic soybean proteins (GIy m Bd 30K, Gly m Bd 28K, and the alpha subunit of conglycinin) from soy protein by using a mutant soybean, Tohoku 124. *Biosci Biotechnol Biochem* **61**:2148–2150 (1997).
31 Takahashi K, Banba H, Kikuchi A, Ito M and Nakamura S, An induced mutant line lacking the α-subunit of β-conglycinin in soybean (*Glycine max* (L) Merril). *Breeding Sci* **44**:65–66 (1994).
32 Kinney AJ, Jung R and Herman EM, Cosuppression of the α-subunits of β-conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum-derived protein bodies. *Plant Cell* **13**:1165–1178 (2001).
33 Takahashi M, Uematsu Y, Kashiwaba K, Yagasaki K, Hajika M, Matsunaga R, *et al*, Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. *Planta* **217**:577–586 (2003).
34 Davies H, Thin layer isoelectric focusing, in *Isoelectric Focusing*, ed. by Arbuthnott JP and Beeley JA. Butterworths, London, pp. 97–113 (1975).
35 Yaklich R, Helm R and Herman EM, Analysis of the distribution of the major soybean allergen in a core collection of *Glycine max* accessions. *Crop Sci* **39**:1444–1447 (1999).
36 Ogawa T, Samoto M and Takahashi K, Development of hypoallergenic soybean products – strategy for the reduction of allergencity of soybean, in *Proceedings of The Third International Soybean Processing and Utilization Conference*, Kourin Publishing Co., Tsukuba, Ibaraki, Japan pp. 165–168 (2000).
37 Joseph LM, Hymowitz T, Schmidt MA and Herman EM, Evaluation of *Glycine* germplasm for nulls of the immunodominant allergen P34/Gly m Bd 30k. *Crop Sci* **46**:1755–1763 (2006).

38 Tsuji H, Bando N, Hiemori M, Yamanishi R, Kimoto M, Nishikawa K, *et al*, Purification of characterization of soybean allergen Gly m Bd 28 K. *Biosci Biotechnol Biochem* **61**:942–947 (1997).

39 Xiang P, Haas EJ, Zeece MG, Markwell J and Sarath G, C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen. *Planta* **220**:56–63 (2004).

40 Tsuji H, Hiemori M, Kimoto M, Yamashita H, Kobatake R, Adachi M, *et al*, Cloning of cDNA encoding a soybean allergen, Gly m Bd 28 K. *Biochim Biophys Acta* **1518**:178–182 (2001).